Investigation of Motion Capture Methods Using Video Cameras: Trial Analysis of Movements during Practice of Playing Musical Instruments

Taku Kawada
Tohoku University, Japan
taku.kawada.d4@tohoku.ac.jp

Yasumasa Yamaguchi
Tohoku University, Japan
yasumasa.yamaguchi.p5@dc.tohoku.ac.jp

Toru Nagahama

Tohoku University, Japan
toru.nagahama.a7@tohoku.ac.jp

Tatsuya Horita
Tohoku University, Japan
horita@horilab.info

In this study, we examined a method for analyzing human motion without using sensors or other devices. Using motion capture technology, which analyzes movements from video images, we developed a program that can analyze movements from video images taken by a common video camera. Using the developed program, we analyzed the movements of a person practicing playing a musical instrument on a trial basis. As a result, upper body movements could be analyzed. Since this research method can analyze movements even from video captured by common video cameras, it is expected to be applied to various fields, including educational situations where it is difficult to use large sensors.

Keywords: Motion Capture, Musical Instruments, Practice, Video Cameras

Introduction

Today, with the development of technology, various devices have become smaller and more powerful, making it easier to measure information (e.g., movement and heart rate) about the learner during learning. For example, Itagaki et al. (2018) attempt to support learning in middle school technology education by using sensors on smartphones to sense planing movements and enable comparison between the model's movements and the learner's own movements. However, there are some problems with motion sensing, such as the need for specialized equipment and the fact that it can only be used in specific places and situations.

On the other hand, due to the increasing sophistication of devices, there are methods of measuring motion that do not use sensors. Kanda et al. (2003) used a motion capture system with 12 infrared cameras with infrared illuminators and markers to analyze human motion. Motion capture required the use of dedicated video cameras. However, inexpensive devices are now being developed that can estimate the human skeleton; Anton et al. (2015) proposed an algorithm that uses an infrared sensor-based device called Kinect to acquire and analyze motion movements. Using their algorithm, they reported that they were able to recognize human posture and motion with high accuracy; Ohashi et al. (2018) proposed an algorithm that can analyze motion using multiple commercially available video cameras, suggesting that it can analyze motion with high accuracy. Castro et al. (2006) also developed and evaluated a system (SOMCAM3D) to measure and analyze movements using commercial video cameras. However, there are some problems with these motion measurements, such as the need for multiple video cameras and dedicated calibration, which make them difficult to use in a simple manner.

In this study, we report on our attempt to develop a method to measure human motion in a simple manner without being restricted by the measurement situation. In this study, we analyzed video images captured by a commercially available video camera for the purpose of measuring simple movements. The subject of measurement and analysis in this study was the movement of practicing playing a musical instrument. Although there are no major movements in playing a musical instrument, we thought that if we could measure movements during playing a musical instrument, we could apply it to the analysis of various movements, because movements change depending on the musical instrument and correct posture is recommended in playing a musical instrument.

Research Design & Methods

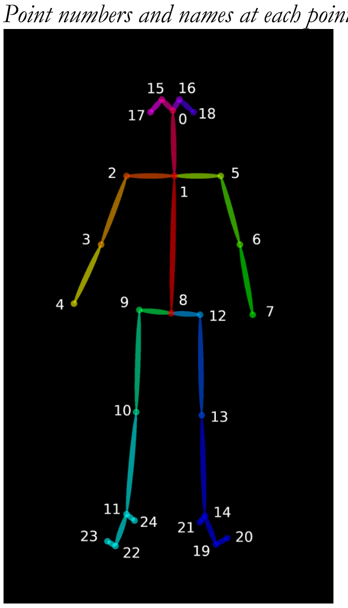
A program for motion analysis was developed to analyze the movements of a performer while playing a musical instrument. In this study, we developed a program using "openpose," which estimates the human skeleton by machine learning using python. The advantages of "openpose" include its relatively flexible configuration and its ability to estimate the skeleton of each person even when multiple people are in the video.

The analysis program estimates the human skeleton in the video for each frame of the video and returns the coordinates of the starting point of the skeleton in the video file. The analyzed coordinates are written to a file for each frame. Table 1 shows the points that are exported to a file. The coordinates of each point are estimated for each frame, and the confidence level of each estimated point is returned as a numerical value.

The participants in the experiment were two Japanese adults (participant A and participant B) who had a certain level of experience playing musical instruments (enough to participate in a competition) and had the opportunity to play musical instruments on a regular basis. After explaining the details of the experiment in writing and orally, the participants signed a consent form. The experimental task was to have the participants play one octave with a fixed rhythm (six beats, two rests) using the instruments they normally play. The participants were then instructed to "play with attention to rhythm and pitch.

During the performance, a large screen was set up about 3 meters in front of the experimental collaborators, and a metronome that presented the rhythm was displayed on the screen. In addition, two conditions were set up: one in which a tuner was placed on a music stand to provide feedback on the pitches played, and the other in which a spectacle-type tuner was used. A video camera was used to capture the upper body of the experiment collaborator while playing the instrument from two locations: (1) the front right side and (2) the left side of the instrument, and the captured video was used for motion analysis. Two cameras were used in this study to examine differences in measurable points depending on the orientation of the video camera.

Table 1
Point numbers and names at each point of the skeleton



Point No	Point X	Point Y
P0	Nose_x	Nose_y
P1	Neck_x	Neck_y
P2	RShoulder_x	RShoulder_y
P3	RElbow_x	RElbow_y
P4	RWrist_x	RWrist_y
P5	LShoulder_x	LShoulder_y
P6	LElbow_x	LElbow_y
P7	LWrist_x	LWrist_y
P8	MidHip_x	MidHip_y
P9	RHip_x	RHip_y
P10	RKnee_x	RKnee_y
P11	RAnkle_x	RAnkle_y
P12	LHip_x	LHip_y
P13	LKnee_x	LKnee_y
P14	LAnkle_x	LAnkle_y
P15	REye_x	REye_y
P16	LEye_x	LEye_y
P17	REar_x	REar_y
P18	LEar_x	LEar_y
P19	LBigToe_x	LBigToe_y
P20	LSmallToe_x	LSmallToe_y
P21	LHeel_x	LHeel_y
P22	RBigToe_x	RBigToe_y
P23	RSmallToe_x	RSmallToe_y
P24	RHeel_x	RHeel_y

Results

The video of the instrument performance was analyzed by the analysis program. The skeleton estimated by openpose and superimposed on the video is shown in Figure 1. The coordinate data of the points with high confidence output by openpose are graphed in Figure 2.

The analysis results of the video taken from the right front are described. When the video taken from the right front was analyzed, most of the points on the upper body had a high level of reliability and could be used for analysis. However, there were cases where the body was hidden by a musical instrument, as in the case of Cooperator A, or where the hands were hidden, as in the case of Cooperator B. In such cases, the reliability was low. In such cases, the reliability is low and it is difficult to use them for analysis.

The following describes the analysis results of the image taken from the left side. When the video taken from the left side was analyzed, the reliability was always high for the nose, left eye, left ear, neck, left shoulder, left elbow, and left wrist, and in some cases, the reliability was also high for the right elbow and right wrist, making it possible to use them for motion analysis. On the other hand, because the images were taken from the left side, the right eye, ear, and shoulder had low reliability and were difficult to use for motion analysis.

Figure 1

Image of the estimated skeleton superimposed on the filmed image.

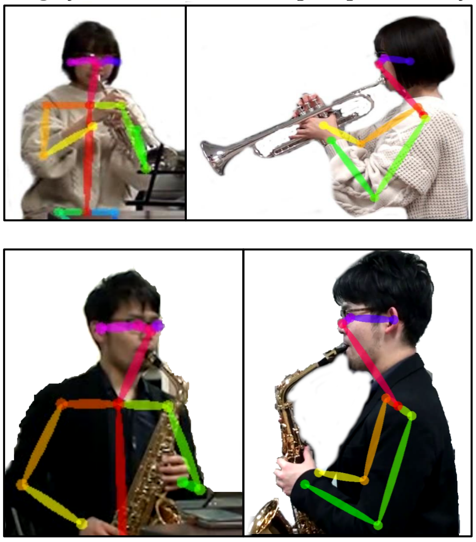
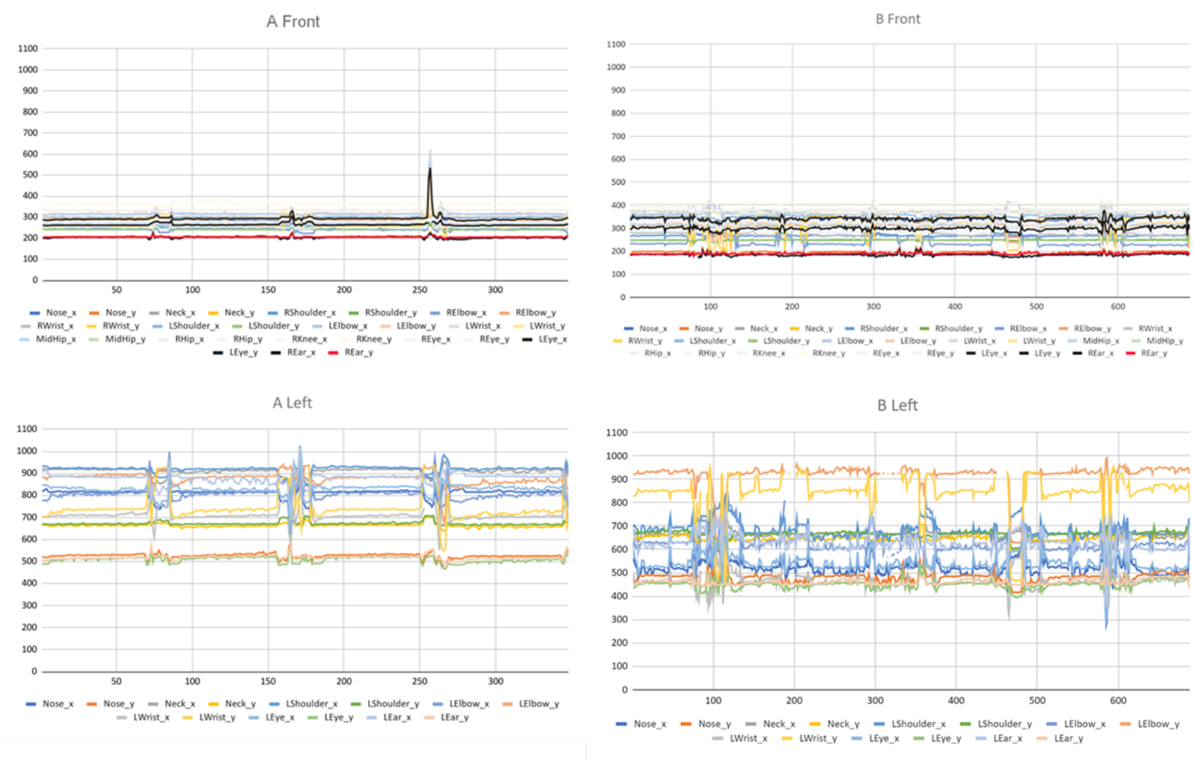


Figure 2

Coordinates at the points of high confidence for participant A and B (top: front, bottom: left side)



Discussion

In this study, we attempted a method to measure human motions in a simple way using motion capture technology. It can be said that the motion of a musical instrument player could be sufficiently estimated even with video captured by a common video camera. In previous studies, motion was analyzed using data captured by large equipment or a large number of video cameras. In this study, however, it is now possible to analyze motion with a single video camera with a certain degree of accuracy by considering the direction and location of the video camera. However, there are some points that cannot be estimated from this study, depending on the shooting location and the type of instrument, and it is necessary to consider the location and orientation of the shooting.

With regard to musical instrument playing, which was the subject of motion analysis in this study, it is thought that it was possible to quantitatively evaluate approximate facial movements, left-right body movements, and shoulder and arm movements from the video shot from the right front side. It is also considered that from the video taken from the left side, it was possible to quantitatively evaluate the vertical movement of the face and posture. The analysis method used in this study demonstrated the possibility of evaluating movements in playing a musical instrument simply by capturing the performance with a video camera. Considering that correct posture is recommended in playing a musical instrument, especially for beginners, it is possible to quantitatively evaluate the degree to which the posture during playing was correct using this method, and it is possible to evaluate practice from a new perspective in musical instrument playing practice. This is expected to provide an opportunity to evaluate and improve performance in terms of posture, which has conventionally been difficult to evaluate in the performance of musical instruments. In addition, since this method can evaluate posture and movement simply by taking pictures with a general video camera, practice can be continued even when the instructor is not available, which is expected to improve the efficiency of individual instrumentalists' practice and reduce the instructor's workload. Furthermore, this research method is capable of time-series analysis of not only movements at a single point in time but also movements during performance, and visualization of movements during performance is expected to serve as an objective indicator of improvement.

Conclusion

In this study, we attempted a method for analyzing performance actions in musical instrument performances. As a result, we were able to analyze the playing motion from video images captured by a general video camera. On the other hand, it is necessary to consider the shooting position and direction of the video camera in the future. Although musical instrument performance was used as the target movement for analysis this time, we would like to apply this method to quantitative evaluation of movements in educational situations in the future. In the field of education, especially in areas where movement is important, such as physical education and technology, it may become possible to qualitatively visualize what has been taught as sensation and provide quantitative advice for improvement. It may also be possible to estimate the state of stress from learners' movements in general classroom situations (Hackford, 2019). In the future, this research will be further developed to develop a motion analysis package that anyone can use more easily and put into practice, including in educational settings.

References

Antón, D., Goni, A., & Illarramendi, A. (2015). Exercise recognition for Kinect-based telerehabilitation. Methods of information in medicine, 54(02), 145-155.

Castro, J. G., Medina-Carnicer, R., & Galisteo, A. M. (2006). Design and evaluation of a new three-dimensional motion capture system based on video. Gait & posture, 24(1), 126-129.

Hackford, J., Mackey, A., & Broadbent, E. (2019). The effects of walking posture on affective and physiological states during stress. Journal of behavior therapy and experimental psychiatry, 62, 80-87.

Itagaki, S., Ando, A., Abiko, H., & Horita, T. (2018). Development and Assessment of a Class System to Support Learning of Planing Motion. Educational technology research, 40(1), 121-132.

Ohashi, T., Ikegami, Y., Yamamoto, K., Takano, W., & Nakamura, Y. (2018). Video motion capture from the part confidence maps of multi-camera images by spatiotemporal filtering using the human skeletal model. In 2018 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS) (pp. 4226-4231). IEEE.